ISSN 1842-6298 (electronic), 1843-7265 (print)
Volume **4** (2009), 155 – 167

ON THE LIU AND ALMOST UNBIASED LIU ESTIMATORS IN THE PRESENCE OF MULTICOLLINEARITY WITH HETEROSCEDASTIC OR CORRELATED ERRORS

M. I. Alheety and B. M. Golam Kibria

Abstract. This paper introduces a new biased estimator, namely, almost unbiased Liu estimator (AULE) of β for the multiple linear regression model with heteroscedastics and/or correlated errors and suffers from the problem of multicollinearity. The properties of the proposed estimator is discussed and the performance over the generalized least squares (GLS) estimator, ordinary ridge regression (ORR) estimator (Trenkler [20]), and Liu estimator (LE) (Kaçiranlar [10]) in terms of matrix mean square error criterion are investigated. The optimal values of d for Liu and almost unbiased Liu estimators have been obtained. Finally, a simulation study has been conducted which indicated that under certain conditions on d, the proposed estimator performed well compared to GLS, ORR and LE estimators.

1 Introduction

Consider the following multiple linear regression model

$$Y = X\beta + \epsilon, \tag{1.1}$$

where Y is an $n \times 1$ vector of observations, X is an $n \times p$ matrix, β is an $p \times 1$ vector of unknown parameters, and ϵ is an $n \times 1$ vector of non observable errors with $E(\epsilon) = 0$ and $Cov(\epsilon) = \sigma^2 I_n$. The most common method used for estimating the regression coefficients in (1.1) is the ordinary least squares (OLS) method which is defined as

$$\hat{\beta} = (X'X)^{-1}X'Y. \tag{1.2}$$

Both the OLS estimator and its covariance matrix heavily depend on the characteristics of the $X'X$ matrix. If XX is ill-conditioned, i.e. the column vectors of X are linearly dependent, the OLS estimators are sensitive to a number of errors. For example, some of the regression coefficients may be statistically insignificant or have the wrong

2000 Mathematics Subject Classification: 62J05; 62F10.
Keywords: Bias, Linear Model; Liu Estimator; MSE; Multicollinearity; OLS; Simulation.

sign, and they may result in wide confidence intervals for individual parameters. With ill-conditioned $X'X$ matrix, it is difficult to make valid statistical inferences about the regression parameters. One of the most popular estimator dealing with multicollinearity is the ordinary ridge regression (ORR) estimator proposed by Hoerl and Kennard [7, 8] and is defined as

$$\tilde{\beta}_k = [X'X + kI_p]^{-1}X'Y = [I_p + k(X'X)^{-1}]^{-1}\hat{\beta}.$$

Both of the Liu estimator $\hat{\beta}_d$ (LE) and the generalized Liu estimator $\hat{\beta}_D$ are defined for each parameter $d \in (-\infty, \infty)$, (see Kaciranlar et al., [9]) as follows

$$\begin{aligned} \hat{\beta}_d &= (X'X + I)^{-1}(X'X + dI)\hat{\beta} \\ \hat{\beta}_D &= (X'X + I)^{-1}(X'X + D)\hat{\beta}, \end{aligned} \tag{1.3}$$

where $D = diag(d_i)$ is a diagonal matrix of the biasing parameter and $d_i \in (-\infty, \infty)$, $i = 1, 2, \cdots, p$ (Akdeniz and Kaciranlar [2]). The advantage of the LE over the ORR is that the LE is a linear function of d, so it is easy to choose d than to choose k in the ORR estimator. Since $X'X$ is symmetric, there exists a $p \times p$ orthogonal matrix P such that $P'X'XP = \Lambda$, Λ is a $p \times p$ diagonal matrix where diagonal elements $\lambda_1, \cdots, \lambda_p$ are the eigenvalues of $X'X$ and $\lambda_1 > \lambda_2 > \cdots > \lambda_p$. So, model (1.1) can be written in the canonical form as:

$$Y = Z\alpha + \epsilon, \tag{1.4}$$

where $Z = XP$ and $\alpha = P'\beta$. The OLS and Liu estimators for (1.4) are respectively

$$\hat{\alpha} = \Lambda^{-1}Z'Y \tag{1.5}$$

and

$$\hat{\alpha}_d = (\Lambda + I)^{-1}(\Lambda + dI)\hat{\alpha}. \tag{1.6}$$

Since $\hat{\alpha}_d$ is a biased estimator, one can apply the jackknife procedure to reduce the bias as proposed by Quenouille [18]. Akdeniz and Kaciranlar [2] proposed the almost unbiased generalized Liu estimator

$$\hat{\alpha}_D = [I - (\Lambda + I)^{-2}(I - D)^2]\hat{\alpha}. \tag{1.7}$$

In model (1.1) we assumed that $Cov(\epsilon) = \sigma^2 I$ which is called homoscedasticity i.e. $Var(\epsilon_i) = \sigma^2$, for $i = 1, \cdots, n$ and uncorrelated i.e. $Cov(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$. Now we made a broader assumption of unequal error variances i.e. heteroscedasticity, that is,

$$E(\epsilon) = 0, \quad Cov(\epsilon) = \sigma^2 V. \tag{1.8}$$

Since $\sigma^2 V$ is the variance-covariance matrix of the errors, V must be positive definite (p.d.), so there exist an $n \times n$ symmetric matrix T, such that $T'T = V$ so that the model (1.1) can be written as

$$T^{-1}Y = T^{-1}XB + T^{-1}\epsilon.$$

Let $Y_* = T^{-1}Y, X_* = T^{-1}X$ and $\epsilon_* = T^{-1}\epsilon$ then $E(\epsilon_*) = 0$ and $Cov(\epsilon_*) = \sigma^2 I$. Therefore the transformed model

$$Y_* = X_*\beta + \epsilon_* \tag{1.9}$$

satisfies the assumption of error $\epsilon_* \sim N(0, \sigma^2 I)$. So the OLS estimator for the model (1.9) is

$$\tilde{\beta} = (X_*'X_*)^{-1}X_*'Y_* = (X'V^{-1}X)^{-1}X'V^{-1}Y \tag{1.10}$$

which is called the generalized least squares (GLS) estimator of β . The GLS estimator is the best linear unbiased estimator of β where $Cov(\tilde{\beta}) = \sigma^2(X'V^{-1}X)^{-1}$. Since the rank of X_* is equal to that of X, then the multicollinearity still also effects the GLS estimator. Trenkler [20] proposed the ridge estimator of β as:

$$\hat{\beta}_k = (X'V^{-1}X + kI)^{-1}X'V^{-1}.$$

Because $\hat{\beta}_k$ is a biased estimator, Özkale [17] proposed a jackknife ridge estimator to reduce the bias of $\hat{\beta}_k$. The organization of the paper is as follows. We propose the almost unbiased Liu estimator (AULE) in Section 2. The performance of AUL estimator compare with other estimators with respect to the matrix mean square error (MSE) and scalar mean square error (mse) are given in Section 3. A simulation study has been conducted in Section 4. Finally some concluding remarks are given in Section 5.

2 The Almost Unbiased Liu Estimators

As we mentioned in Section 1, the structure of LE is obtained by combining the ridge philosophy with the Stein estimator [19]. Kaciranlar [10] introduced the Liu estimator under the general linear model as follows

$$\tilde{\beta}_d = (X'V^{-1}X + I)^{-1}(X'V^{-1}X + dI)\tilde{\beta}. \tag{2.1}$$

By using the canonical form, we may rewrite (2.1) as follows:

$$\tilde{\alpha}_d = (\Gamma + I)^{-1}(\Gamma + dI)\tilde{\alpha}, \tag{2.2}$$

where

$$\tilde{\alpha} = (W'W)^{-1}W'Y_* = \Gamma^{-1}Q'X'V^{-1}Y, \tag{2.3}$$

where $W = X_*Q = T^{-1}XQ$, Q is an orthogonal matrix of $X'V^{-1}X$ which those columns are the eigenvector of the matrix $X'V^{-1}X$ and

$$W'W = \Gamma = diag\{\gamma_1, \gamma_2, \cdots, \gamma_p\}$$

is the diagonal matrix which those elements are the eigenvalues of the matrix $X'V^{-1}X$.

It is clear that $\tilde{\alpha}_d$ is biased estimator (See Kaciranlar [10])

$$Bias(\tilde{\alpha}_d) = -(1-d)(\Gamma+I)^{-1}\alpha.$$

The variance-covariance matrix of $\tilde{\alpha}_d$ is given as follows:

$$Cov(\tilde{\alpha}_d) = \sigma^2\Gamma^{-1}(\Gamma+I)^{-2}(\Gamma+dI)^2.$$

In order to compare the performance of any estimator with others, a criterion for measuring the goodness of an estimator is required. For this purpose, the mean square error (MSE) criteria is used to measure the goodness of an estimator. We note that for any estimator β^* of β, its MSE is defined as

$$MSE(\beta^*) = E(\beta^*-\beta)(\beta^*-\beta)' = Cov(\beta^*) + Bias(\beta^*)Bias(\beta^*)'.$$

Therefore the MSE of $\tilde{\alpha}_k$ is given as follows:

$$MSE(\tilde{\alpha}_k) = \sigma^2\Gamma(\Gamma+kI)^{-2} + k^2(\Gamma+kI)^{-1}\alpha\alpha'(\Gamma+kI)^{-1}.$$

Also, the MSE of $\tilde{\alpha}_d$ is obtained as:

$$MSE(\tilde{\alpha}_d) = \sigma^2\Gamma^{-1}(\Gamma+I)^{-2}(\Gamma+dI)^2 + (d-1)^2(\Gamma+I)^{-1}\alpha\alpha'(\Gamma+I)^{-1} \tag{2.4}$$

and the scalar mean square error (mse) is obtained as follows

$$mse(\tilde{\alpha}_d) = tr(MSE(\tilde{\alpha}_d)) = \sum_{i=1}^{p}\frac{\sigma^2(\gamma_i+d)^2 + (d-1)^2\gamma_i\alpha_i^2}{\gamma_i(\gamma_i+1)^2}. \tag{2.5}$$

Since $\tilde{\alpha}_d$ is a biased estimator, one can apply the jackknife procedure to reduce the bias of $\tilde{\alpha}_d$ which is due to Quenouille [18]. We may rewrite (2.2) as follows:

$$\begin{aligned}\tilde{\alpha}_d &= \tilde{\alpha} - (1-d)(\Gamma+I)^{-1}\tilde{\alpha},\\ Bias(\tilde{\alpha}_d) &= -(1-d)(\Gamma+I)^{-1}\alpha.\end{aligned}$$

Thus, following Kadiyala [11] and Ohtani [16], we have

$$\begin{aligned}\tilde{\alpha}_d^* &= [I+(\Gamma+I)^{-1}(1-d)]\tilde{\alpha}_d \\ &= [I-(\Gamma+I)^{-2}(1-d)^2]\tilde{\alpha}\end{aligned} \tag{2.6}$$

which we called AUL estimator. The form in (2.6) seems to be the same with the jackknife Liu estimator that Akdeniz and Kaçiranlar [2] proposed, but in (2.6), $W'W=\Gamma$ is the diagonal matrix of the eigenvalues of $X'V^{-1}X$ whereas in Akdeniz and Kaciranlar [2] $Z'Z=\Lambda$ is the diagonal matrix of the eigenvalues of $X'X$. The bias and the variance-covariance matrix of $\tilde{\alpha}_d$ are given as follows:

$$\begin{aligned}Bias(\tilde{\alpha}_d) &= -(1-d)^2(\Gamma+I)^{-2}\alpha \\ Cov(\tilde{\alpha}_d^*) &= \sigma^2[I-(1-d)^2(\Gamma+I)^{-2}]\Gamma^{-1}[I-(1-d)^2(\Gamma+I)^{-2}].\end{aligned} \tag{2.7}$$

Therefore, the MSE and mse of $\tilde{\alpha}_d^*$ are

$$\begin{aligned} & MSE(\tilde{\alpha}_d^*) = \\ = \quad & \sigma^2[I-(1-d)^2(\Gamma+I)^{-2}]\Gamma^{-1}[I-(1-d)^2(\Gamma+I)^{-2}]+ \\ & \qquad\qquad\qquad + (1-d)^4(\Gamma+I)^{-2}\alpha\alpha'(\Gamma+I)^{-2} \end{aligned} \tag{2.8}$$

and

$$mse(\tilde{\alpha}_d^*) = \sum_{i=1}^{p} \frac{\sigma^2(\gamma_i+1)^4\left(1-\frac{(1-d)^2}{(\gamma_i+1)^2}\right)^2 + \gamma_i(1-d)^4\alpha_i^2}{\gamma_i(\gamma_i+1)^4}. \tag{2.9}$$

3 Performance of the Estimators

This section compare the performance of the estimators with the smaller MSE criteria. For the sake of convenience, we define some Lemmas which are presented below.

Lemma 1. *Let M be a p.d. matrix , and a be a vector, then $M - aa'$ is non negative definite matrix (n.n.d.) if and only if $a'Ma \leq 1$.*

Proof. See Farebrother [4]. □

Lemma 2. *Let $\hat{\beta}_j = A_jY, j = 1,2$ be two linear estimators of β. Suppose that $D = Cov(\hat{\beta}_1) - Cov(\hat{\beta}_2)$ is p.d. then $\Delta = MSE(\hat{\beta}_1) - MSE(\hat{\beta}_2)$ is n.n.d. if and only if $b_2'(D + b_1b_1')^{-1}b_2 \leq 1$, where b_j denotes the bias vector of $\hat{\beta}_j$.*

Proof. See Trenkler and Toutenburg [21]. □

3.1 Comparison between AUL Estimator and GLS Estimator

The difference MSE between the AUL and GLS estimators is given as follows

$$\Delta_1 = MSE(\tilde{\alpha}) - MSE(\tilde{\alpha}_d^*) = \sigma^2 D_1 - b_2b_2',$$

where $D_1 = \Gamma^{-1} - \Gamma^{-1}[I-(1-d)^2(\Gamma+I)^{-2}][I-(1-d)^2(\Gamma+I)^{-2}]$ and $b_2 = -(1-d)^2(\Gamma+I)^{-2}\alpha$.

In the following theorem, we will give the necessary and sufficient conditions for $\tilde{\alpha}_d^*$ to be superior to the $\tilde{\alpha}$.

Theorem 3. *Under the linear regression model with heteroscedastic and/or correlated errors, the $\tilde{\alpha}_d^*$ is superior to the $\tilde{\alpha}$ in the MSE sense, namely, Δ_1 if and only if*

$$b_2'D_1^{-1}b_2 \leq \sigma^2.$$

Proof. It is clear that when $0 < d < 1$, then $I - [I - (I - dI)^2(\Gamma + I)^{-2}]^2$ is p.d. and that means D_1 is p.d. Therefore using Lemma 1, Δ_1 is n.n.d. if and only if $b_2' D_1^{-1} b_2 \leq \sigma^2$ and by using an explicit form, Δ_1 is n.n.d. if and only if

$$(1-d)^2 \alpha' \Gamma \left[2(\Gamma + I)^2 - (I - d)^2 I\right]^{-1} \alpha < \sigma^2.$$

The proof is completed. □

3.2 Comparison between AUL Estimator and Liu Estimator

Consider

$$\Delta_2 = MSE(\tilde{\alpha}_d) - MSE(\tilde{\alpha}_d^*) = \sigma^2 D_2 + b_1 b_1' - b_2 b_2',$$

where $D_2 = \Gamma^{-1}(\Gamma + I)^{-2}(\Gamma + dI)^2 - \Gamma^{-1}[I - (1-d)^2(\Gamma + I)^{-2}]^2$ and $b_1 = -(1 - d)(\Gamma + I)^{-1}\alpha$. The following theorem gives the necessary and sufficient conditions for $\tilde{\alpha}_d^*$ to be superior to the $\tilde{\alpha}_d$.

Theorem 4. *Under the linear regression model with heteroscedastic and/or correlated errors, when $1 < d < 2\gamma_i + 3$, the $\tilde{\alpha}_d^*$ is superior to the $\tilde{\alpha}_d$ in the MSE sense, namely Δ_2, if and only if $b_2'(\sigma^2 D_2 + b_1 b_1')^{-1} b_2 \leq 1$.*

Proof. D_2 is p.d. if and only if

$$\begin{aligned} &\frac{(\gamma_i + 1)^2(\gamma_i + d)^2 - \left[(\gamma_i + 1)^2 - (\gamma_i - d + 2)^2\right]}{\gamma_i(\gamma_i + 1)^4} \\ = &\frac{(\gamma_i + d)^2(d-1)(2\gamma_i - d + 3)}{\gamma_i(\gamma_i + 1)^4} > 0. \end{aligned}$$

Since $(\gamma_i + d)^2 > 0$ for any value of d, we have to consider the following option to show that D_2 is p.d.

- If $d > 1$, then $(2\gamma_i - d + 3)$ must be greater than zero, i.e.

$$(2\gamma_i - d + 3) > 0 \Rightarrow d < 2\gamma_i + 3.$$

We should find d that makes $(2\gamma_i - d + 3) > 0$. This is will happen when $d < 2\gamma_i + 3$. Therefore, when $1 < d < 2\gamma_i + 3$, D_2 is p.d. By applying Lemma 2, Δ_2 is n.n.d. if and only if

$$b_2'(\sigma^2 D_2 + b_1 b_1')^{-1} b_2 \leq 1.$$

By using an explicit form, Δ_2 is n.n.d. if and only if

$$\begin{aligned} (1-d)^4 \alpha'(\Gamma + I)^{-2} &\left[\sigma^2 \Gamma^{-1}(\Gamma + I)^{-2}(\Gamma + dI)^2 - \Gamma^{-1}[I - (1-d)^2(\Gamma + I)^{-2}]^2 \right. \\ &\left. + (1-d)^2(\Gamma + I)^{-1}\alpha\alpha'(\Gamma + I)^{-1}\right](\Gamma + I)^{-2}\alpha \leq 1. \end{aligned}$$

The proof is completed. □

3.3 Comparison between AUL Estimator and ORR Estimator

Consider

$$\begin{aligned}\Delta_3 &= MSE(\tilde{\alpha}_k) - MSE(\tilde{\alpha}_d) = \sigma^2 D_3 + b_3 b_3' - b_2 b_2' \\ \Delta_4 &= MSE(\tilde{\alpha}_d) - MSE(\tilde{\alpha}_k) = \sigma^2 D_4 + b_2 b_2' - b_3 b_3'\end{aligned}$$

where

$$\begin{aligned}D_3 &= (\Gamma + kI)^{-1}\Gamma(\Gamma + kI)^{-1} - (\Gamma + I)^{-2}(\Gamma + dI)^2\Gamma^{-1} \\ D_4 &= (\Gamma + I)^{-2}(\Gamma + dI)^2\Gamma^{-1} - (\Gamma + kI)^{-1}\Gamma(\Gamma + kI)^{-1} \\ b_3 &= -k(\Gamma + kI)^{-1}\alpha\end{aligned}$$

Let us assume that k is fixed. We need to show that D_3 is p.d.,

$$D_3 = diag\left\{\frac{\gamma_i}{(\gamma_i + k)^2} - \frac{(\gamma_i + d)^2}{\gamma_i(\gamma_i + 1)^2}\right\}_{i=1}^{p} \text{ is p.d. if}$$

$$\left\{\frac{\gamma_i^2(\gamma + 1)^2 - (\gamma_i + k)^2(\gamma_i + d)^2}{\gamma_i(\gamma + 1)^2(\gamma_i + k)^2}\right\} > 0$$

$$\Leftrightarrow \qquad d < \frac{(1-k)\gamma_i}{(\gamma_i + k)}.$$

Therefore, D_3 is p.d. for $0 < d < d^* = \frac{(1-k)\gamma_i}{(\gamma_i + k)}$. After applying Lemma 2, we can state the following theorem.

Theorem 5. *Under the linear regression model with heteroscedastic and/or correlated errors , then*

1. *For* $0 < d < d^*$, Δ_3 *is n.n.d. if and only if*

$$b_2'(\sigma^2 D_3 + b_3 b_3')^{-1} b_2 < 1$$

and if we use an explicit form, then Δ_3 *is n.n.d. if and only if*

$$\begin{aligned}&(1-d)^4\alpha'[\sigma^2(\Gamma + I)^2\Gamma(\Gamma + kI)^{-2}(\Gamma + I)^2 - (\Gamma + dI)^2\Gamma^{-1}(\Gamma + I)^2 + \\ &+ k^2(\Gamma + I)^2(\Gamma + kI)^{-1}\alpha\alpha'(\Gamma + kI)^{-1}(\Gamma + I)^2]^{-1}\alpha < 1.\end{aligned}$$

2. *For* $0 < d^* < d < 1$, Δ_4 *is n.n.d. if and only if*

$$b_3'(\sigma^2 D_4 + b_2' b_2)^{-1} b_3 < 1$$

and if we use an explicit form, then Δ_4 *is n.n.d. if and only if*

$$\begin{aligned}&k^2\alpha'[\sigma^2(\Gamma + kI)(\Gamma + I)^{-2}(\Gamma + dI)^2\Gamma^{-1}(\Gamma + kI) - \Gamma + \\ &(1-d)^4(\Gamma + kI)(\Gamma + I)^{-2}\alpha\alpha'(\Gamma + I)^{-2}(\Gamma + kI)]^{-1}\alpha < 1.\end{aligned}$$

Let us now assumed that d is fixed. By using same approach in D_3, D_4 is p.d. for $k > k^* = \frac{(1-d)\gamma_i}{(\gamma_i+d)}$. After applying Lemma 2 we have the following theorem

Theorem 6. *Under the linear regression model with heteroscedastic and/or correlated errors, then*

1. *For $k > k^*$, Δ_4 is n.n.d. if and only if*

$$b_3'(D_4 + b_2 b_2')^{-1} b_3 < 1.$$

2. *For $k < k^*$, Δ_3 is n.n.d. if and only if*

$$b_2'(D_3 + b_3 b_3')^{-1} b_2 < 1.$$

The explicit forms of Theorem 6 are same as the explicit forms of Δ_3 and Δ_4 in Theorem 5. We can find the optimal value of d by minimizing the MSE of Liu estimator and AUL estimator in the linear model with heteroscedastic and/or correlated errors as follows:

$$d = \frac{\sum \frac{\gamma_i(\alpha^2-\sigma^2)}{\gamma_i(\gamma_i+1)^2}}{\sum \frac{(\sigma^2+\gamma_i\alpha_i^2)}{\gamma_i(\gamma_i+1)^2}} \tag{3.1}$$

and

$$d_{AULE} = 1 - \sqrt{\frac{\sum_{i=1}^{p} \frac{\sigma^2}{\gamma_i(\gamma_i+1)^2}}{\sum_{i=1}^{p} \frac{\sigma^2+\gamma_i\alpha_i^2}{\gamma_i(\gamma_i+1)^4}}} \tag{3.2}$$

respectively.

Since d depends on the unknown parameters (α, σ^2), we replace them by their unbiased estimators (GLS) and the estimated values are

$$\hat{d} = \frac{\sum\limits_{i=1}^{p} \frac{\gamma_i(\hat{\alpha}_i^2-\hat{\sigma}^2)}{\gamma_i(\gamma_i+1)^2}}{\sum\limits_{i=1}^{p} \frac{(\hat{\sigma}^2+\gamma_i\hat{\alpha}_i^2)}{\gamma_i(\gamma_i+1)^2}}$$

and

$$\hat{d}_{AULE} = 1 - \sqrt{\frac{\sum_{i=1}^{p} \frac{\hat{\sigma}^2}{\gamma_i(\gamma_i+1)^2}}{\sum_{i=1}^{p} \frac{\hat{\sigma}^2+\gamma_i\hat{\alpha}_i^2}{\gamma_i(\gamma_i+1)^4}}}$$

respectively.

We can note from our theorems that the comparison results depend on the unknown parameters β and σ^2. In consequence of that, we cannot exclude that our results obtained in the theorems will be held and the results may be changeable.

So, we replace them (β and σ^2) by their unbiased estimators. Since V is rarely known, the estimate of V can be used. Trenkler [20] gave some estimates of V as

$$V = (v_{ij})\,, \qquad v_{ij} = \rho^{|i-j|}, \qquad i,j = 1,2,...,n \tag{3.3}$$

and

$$V = \frac{1}{1+\rho^2} = \begin{pmatrix} 1+\rho^2 & \rho & 0 & \cdots & 0 \\ \rho & 1+\rho^2 & \rho & \cdots & 0 \\ 0 & \rho & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1+\rho^2 & \rho \\ 0 & 0 & \cdots & \rho & 1+\rho^2 \end{pmatrix}.$$

Firinguetti [5] introduced another estimate of V matrix as follows

$$V = \frac{1}{1-\rho^2} = \begin{pmatrix} 1 & \rho & \cdots & \rho^{n-1} \\ \rho & 1 & \cdots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \cdots & 1 \end{pmatrix}. \tag{3.4}$$

There are other estimates of V matrix are given by other researchers like Bayhan and Bayhan [3].

4 The Monte Carlo Simulation

This section conducted a simulation study to compare the performance of the AULE with other estimators. To achieve different degrees of collinearity, following McDonald and Galarneau [14], Gibbons [6] and Kibria [12], the explanatory variables are generated by using the following equation.

$$x_{ij} = (1-\gamma^2)^{1/2} z_{ij} + \gamma z_{ip}, i = 1,\cdots,n, j = 1,2,\cdots,p,$$

where z_{ij} are independent standard normal pseudo-random numbers , $p = 5$ is the number of the explanatory variables, $n = 50, 100$ and 150 and γ is specified so that the correlation between any two explanatory variables is given by γ^2. Three different sets of correlation are considered according to the value of $\gamma = 0.85$ and 0.95. Also the explanatory variables are standardized so that $X'X$ will be in correlation form.

The estimated V matrix that given in (3.3) is used in this simulation study where four values of ρ are given as 0.1, 0.4, 0.7, 0.9. According to Kibria [12], Alheety and Gore [1] and Muniz and Kibria [15], we consider the coefficient vector that corresponded to the largest eigenvalue of $X'V^{-1}X$ matrix. The n observations for the dependent variable are determined by the following equation:

$$y_i = \beta_0 + \beta_1 x_{i1} \ + \ \beta_2 x_{i2} \ + \ \beta_3 x_{i3} \ + \ \cdots \ + \ \beta_p x_{ip} + e_i,$$

where e_i are independent normal pseudo-random numbers with mean 0 and variance σ^2. In this study, β_0 is taken to be zero. Three values of σ are given as 0.01, 0.3 and 1. The estimated optimal values of d and d_{AULE} are used for Liu and AUL estimator respectively. The experiment is replicated 2000 times by generating new error terms. The MSEs for the estimators are calculated as follows

$$MSE(\hat{\beta}^*) = \frac{1}{2000} \sum_{r=1}^{2000} (\hat{\beta}^*_{(r)} - \beta)'(\hat{\beta}^*_{(r)} - \beta),$$

where $\hat{\beta}^*$ is any estimator that used in this study for making a comparison.

The simulated MSEs for all estimators are presented in Table 1. It is observed that the proposed AULE estimator performing well compare to others for small σ. It is also noted that the performance of the estimators depend on ρ, γ and the sample size n.

5 Concluding Remarks

A new biased estimator has been proposed for estimating the parameter of the linear regression model with multicollinearity and heteroscedastics and/or correlated errors. The bias and MSE expressions of the estimators are given. The proposed almost unbiased Liu estimator (AULE) is examined against generalized least squares (GLS), ordinary ridge regression (ORR) and Liu estimators in terms of matrix mean square error criterion. The optimal values of d for Liu and almost unbiased Liu estimators are obtained. A simulation study has been made to show the performance of the proposed estimator compared with others. It is observed that under some conditions on d, the proposed estimator performed well compared to others. We belive that the findings of this paper will be useful for the practitioners.

Acknowledgment. The authors are thankful to the editor and referee for their constructive and valuable suggestions and comments that have improved the presentation of the paper greatly.

Table 1: Mean square errors of the GLS, LE, AULE and ORR estimators for different ρ and V estimated using the equation (3.3)

ρ	n	σ	$\gamma = 0.85$				$\gamma = 0.95$			
			GLS	LE	AULE	ORR	GLS	LE	AULE	ORR
0.10	50	0.01	0.2401	0.2401	0.2401	0.2401	0.2901	0.2901	0.2901	0.2901
		0.3	0.2414	0.2457	0.2364	0.2414	0.2925	0.3069	0.2875	0.2925
		1	0.3722	0.3930	0.4280	0.3618	0.5384	0.4563	0.748	0.4612
	100	0.01	0.2539	0.2539	0.2539	0.2539	0.2805	0.2805	0.2805	0.2805
		0.3	0.2546	0.2570	0.2494	0.2546	0.2819	0.2893	0.2751	0.2819
		1	0.331	0.3668	0.3489	0.3297	0.4268	0.4249	0.5037	0.4119
	150	0.01	0.244	0.2440	0.2440	0.2440	0.2868	0.2868	0.2868	0.2868
		0.3	0.2447	0.2463	0.2413	0.2448	0.2879	0.2938	0.2811	0.2879
		1	0.3256	0.3653	0.3318	0.3254	0.3972	0.4086	0.4507	0.3893
0.40	50	0.01	0.2475	0.2475	0.2475	0.2475	0.2827	0.2827	0.2827	0.2827
		0.3	0.2488	0.2519	0.2465	0.2488	0.2848	0.2940	0.2823	0.2849
		1	0.3788	0.3821	0.4292	0.3721	0.5000	0.4417	0.6285	0.4623
	100	0.01	0.2533	0.2533	0.2533	0.2533	0.2844	0.2844	0.2844	0.2844
		0.3	0.254	0.2561	0.2512	0.2541	0.2858	0.2919	0.2827	0.2858
		1	0.336	0.3545	0.3578	0.3348	0.4219	0.4076	0.4893	0.4101
	150	0.01	0.25	0.2500	0.2500	0.2500	0.2898	0.2898	0.2898	0.2898
		0.3	0.2507	0.2520	0.2483	0.2507	0.2907	0.2952	0.2871	0.2907
		1	0.3283	0.3513	0.3398	0.3278	0.3905	0.3895	0.4346	0.3848
0.70	50	0.01	0.2336	0.2336	0.2336	0.2336	0.2909	0.2909	0.2909	0.2909
		0.3	0.2355	0.2379	0.2339	0.2355	0.2936	0.2983	0.2941	0.2936
		1	0.413	0.4026	0.4498	0.4069	0.5810	0.5106	0.6794	0.5465
	100	0.01	0.2411	0.2411	0.2411	0.2411	0.2824	0.2824	0.2824	0.2824
		0.3	0.2421	0.2432	0.2407	0.2421	0.2841	0.2883	0.2835	0.2841
		1	0.3428	0.3481	0.3557	0.3422	0.4546	0.4293	0.5019	0.4449
	150	0.01	0.2364	0.2364	0.2364	0.2364	0.2833	0.2833	0.2833	0.2833
		0.3	0.2373	0.2382	0.2358	0.2373	0.2845	0.2874	0.2835	0.2845
		1	0.3312	0.3391	0.3397	0.3310	0.4030	0.3932	0.4312	0.3998
0.90	50	0.01	0.2387	0.2387	0.2387	0.2387	0.2887	0.2887	0.2887	0.2887
		0.3	0.2403	0.2409	0.2404	0.2403	0.2911	0.2924	0.2914	0.2911
		1	0.4081	0.4000	0.4216	0.406	0.5558	0.5279	0.5886	0.5455
	100	0.01	0.2397	0.2397	0.2397	0.2397	0.2868	0.2868	0.2868	0.2868
		0.3	0.2407	0.2412	0.2406	0.2407	0.2883	0.2895	0.2884	0.2883
		1	0.3478	0.3466	0.3538	0.3475	0.4462	0.435	0.4606	0.4438
	150	0.01	0.2415	0.2415	0.2415	0.2415	0.2871	0.2871	0.2871	0.2871
		0.3	0.2424	0.2428	0.2421	0.2424	0.2882	0.2892	0.2882	0.2882
		1	0.3371	0.3374	0.3411	0.337	0.4054	0.3986	0.4153	0.4043

References

[1] M. I. Alheety and S. D. Gore, *A new estimator in multiple linear regression*, Model Assisted Statistics and Applications, **3** (3) (2008), 187-200.

[2] F. Akdeniz and S. Kaçiranlar, *On the almost unbiased generalized Liu estimator and unbiased estimation of the bias and MSE*, Communications in Statistics-Theory and Methods, **24** (1995), 1789-1797. MR1350171. Zbl 0937.62612.

[3] G. M. Bayhan and M. Bayhan, *Forecasting using autocorrelated errors and multicollinear predictor variables*, Computers and Industrial Engineering, **34** (1998), 413-421.

[4] R. W. Farebrother, *Further results on the mean square error of ridge regression*, Journal of Royal Statistical Society, **B.38** (1976), 284-250. MR0653156(58 #31610). Zbl 0344.62056.

[5] L. Firinguetti, *A simulation study of ridge regression estimators with autocorrelated errors*, Communications in Statistics - simulation and computation, **18** (2) (1989), 673-702. Zbl 0695.62176.

[6] D. G. Gibbons, *A simulation study of some ridge estimators*, Journal of the American Statistical Association, **76** (1981), 131139. Zbl 0452.62055.

[7] A. E. Hoerl and R. W. Kennard, *Ridge Regression : Biased estimation for non-orthogonal problem*, Technometrics, **12** (1970a), 55-67.

[8] A. E. Hoerl and R. W. Kennard, *Ridge Regression : Biased estimation for non-orthogonal problem*, Technometrics, **12** (1970a), 69-82.

[9] S. Kaçiranlar, S. S. F. Akdeniz, G. P. H. Styan and H. J. Werner, *A new biased estimator in linear regression and detailed analysis of the widely-analysed dataset on Portland Cement*, Sankhya B, **61** (1999), 443-459.

[10] S. Kaçiranlar, *Liu estimator in the general linear regression model*, Journal of Applied Statistical Science, **13** (2003), 229-234. MR2038812 (2004k:62170). Zbl 1059.62071.

[11] K. Kadiyala, *A class almost unbiased and efficient estimators of regression coefficients*, Economics Letter, **16** (1984), 293-296.

[12] B. M. G. Kibria, *Performance of some new ridge regression estimators*, *Communications in Statistics-Simulation and Computation*, **32** (2003), 419-435. MR1983342. Zbl 1075.62588.

[13] K. Liu, *A new class of biased estimate in linear regression*, Communications in Statistics-Theory and Methods, **22** (1993), 393-402. MR1212418.

[14] G. C. Mcdonald and D. I. Galarneau, *A monte carlo evaluation of some ridge type estimators*, Journal of the American Statistical Association, **20** (1975), 407-416. Zbl 0319.62049.

[15] G. Muniz and B. M. G. Kibria, *On Some Ridge Regression Estimators: An Empirical Comparisons*, Communications in Statistics - Simulation and Computation, **38** (2009), 621630.

[16] K. Ohtani, *On small sample properties of the almost unbiased generalized ridge estimator*, Communications in Statistics-Theory and Methods, **15** (1986), 1571-1578. MR0845856. Zbl 0611.62079.

[17] M. R. Özkale, *A jackknifed ridge estimator in the linear regression model with heteroscedastic or correlated errors*, Statistics and Probability Letters, **78** (18) (2008), 3159-3169. MR2479473. Zbl pre05380098.

[18] M. Quenouille, *Notes on bias in estimation*, Biometrika, **43** (1956), 3-4, 353-360. MR0081040(18,344f).

[19] C. Stein, *Inadmissibility of the usual estimator for the mean of a multivariate normal distribution*, In : Proceedings of the Third Berkeley Symposium on Mathematics, Statistics and Probability, University of California, Berkeley, **I** (1956), 197-206. MR0084922(18,948c).

[20] G. Trenkler, *On the performance of biased estimators in the linear regression model with correlated or heteroscedastic errors*, Journal of Econometrics, **25** (1984), 179-190. MR0748043(86a:62103).

[21] G. Trenkler and H. Toutenburg, *Mean squared error matrix comparisons between biased estimators An overview of recent results*, Statistical Papers, **31** (1) (1990), 165-179. MR1124154(92h:62104). Zbl 0703.62066.

Mustafa I. Alheety
Department of Mathematics
Alanbar University, Iraq
e-mail: alheety@yahoo.com

B. M. Golam Kibria
Department of Mathematics and Statistics
Florida International University, Miami, FL 33199, USA.
e-mail: kibriag@fiu.edu